Exact reconstruction of sparse signals via nonconvex minimization

Rick Chartrand
Los Alamos National Laboratory
EDICS: DSP-TFSR

***Abstract*— Several authors have shown recently that is possible to reconstruct exactly a sparse signal from fewer linear measurements than would be expected from traditional sampling theory. The methods used involve computing the signal of minimum ℓ^1 norm among those having the given measurements. We show that by replacing the ℓ^1 norm with the ℓ^p norm with $p < 1$, exact reconstruction is possible with substantially fewer measurements. We give many numerical examples in one complex dimension, and larger-scale examples in two real dimensions.**

I. Introduction

There has been much recent research (e.g., [1], [2], [3]) on the subject of reconstruction of sparse signals from a limited number of linear measurements. This topic is known in the literature as *compressed sensing* or *compressive sampling*, and has some overlap with *basis pursuit* [4]. There are many relevant results, but we focus here on those typified by the following example, from [3].

We consider an $M \times N$ measurement matrix Φ such that for $x \in \mathbb{C}^N$, $y = \Phi x$ is the vector of Fourier coefficients of x at M randomly chosen frequencies. Suppose that the *sparsity* of x is K; that is, x has K nonzero elements. This can be stated in terms of the ℓ^0 norm[1] of x: $\|x\|_0 = K$. Then there is a constant C, not depending on K or N, such that whenever $M > CK \log N$, one can reconstruct x exactly, with very high probability, as the solution of the following optimization problem:

$$\min_u \|u\|_0, \quad \text{subject to} \quad \Phi u = y. \tag{1}$$

(The constant C can depend on how small the probability of failure is to be guaranteed to be; it scales as N^{-m} where m depends on C.) Importantly, this result continues to hold if the ℓ^0 norm is replaced by the ℓ^1 norm, resulting in a convex problem:

$$\min_u \|u\|_1, \quad \text{subject to} \quad \Phi u = y. \tag{2}$$

Address: Theoretical Division, T-7, MS B284, Los Alamos National Laboratory, Los Alamos, NM 87544, USA. Phone: +1 +505 667-8093. Fax: +1 +505 665-5757. E-mail: rickc@lanl.gov.

[1]This is a standard abuse of terminology: $\|\cdot\|_0$ is not positive homogeneous, yet is referred to as a norm.

The problem (1) is NP-hard [5], while the convex problem (2) can be solved efficiently (using linear programming, for example). That the former can be replaced by the latter is the reason for the surge in recent interest.

It should be pointed out that the above is not special to Fourier measurements. Similar results hold [1] if the elements of Φ consist of random samples from the standard normal distribution, or if each element is 1 or -1 with equal probability. The key is that with high probability, Φ will be a sampling from a basis that is *incoherent* with the standard basis in $\mathbb{C}^N$. Just as a signal cannot be localized in both time and frequency, there is a lower bound on the combined support size of x and Φx.

It is natural to ask what happens if the ℓ^1 norm is replaced by the ℓ^p norm[2] for some $p \in (0, 1)$. The resulting optimization problem will not be convex, and is described in the literature as "intractable." In this note, we demonstrate that this problem can be solved tractably, if not efficiently, and allows for exact reconstruction of sparse signals with many fewer measurements than when $p = 1$. Our numerical evidence consists of a suite of one-dimensional examples and some larger-scale, two-dimensional examples. The examples are designed for direct comparison with results in [3]. We also include an example illustrating robustness with respect to noise.

II. One-dimensional examples

We begin with a numerical investigation of the circumstances under which the solution to the problem described in the introduction,

$$\min_u \|u\|_p, \quad \text{subject to} \quad \Phi u = \Phi x, \tag{3}$$

equals exactly x. As before, Φ will be a matrix having the action of evaluation of the Fourier transform at a randomly chosen collection of frequencies.

For our experiments, we fixed a signal length of $N = 512$. As in [3], the numbers of measurements M were 8, 16, ..., 128. For each value of K, the sparsities K are each multiple of $M/16$ up to M, rounded to

[2]As with $\|\cdot\|_0$, $\|\cdot\|_p$ is not a norm when $0 < p < 1$, though $\|\cdot\|_p^p$ satisfies the triangle inequality and induces a metric.

the nearest integer. The values of p were 1, 0.95, 0.75, and 0.5. (Smaller values of p can be used, but only with more careful supervision of parameters than that afforded by the automated approach described below.) The experiment was repeated 10 times for each value of M and K. For each repetition, a spike-train signal $x = \sum_{k=1}^{K} z_k e_k$ was constructed by randomly choosing K elements of the standard basis $\{e_1, e_2, \ldots, e_N\}$ of $\mathbb{C}^N$, then randomly choosing the real and imaginary parts of each z_k from the standard normal distribution. Conceptually, the matrix Φ consisted of a random selection of M rows from the discrete Fourier transform (DFT) matrix. In actual fact, instead of multiplication by a matrix Φ, the fast Fourier transform would be computed and the corresponding M elements selected.

The solution to problem (3) was computed by gradient descent with projection. To avoid division by zero, $|u|$ was replaced by $\sqrt{|u|^2 + \epsilon^2}$, where ϵ was decremented by 5% every 100 iterations, beginning with 0.01. Every 100 iterations, an exact line search was performed, then this stepsize used until the next line search. The projection onto $\Phi u = \Phi x$ was performed every iteration, by computing the DFT of u, setting the previously-chosen M frequency coefficients to those of x, then computing the inverse DFT. This projection was also incorporated into the line search: the stepsize was chosen to minimize the ℓ^p norm of the projection along the ray in the steepest descent direction. The iteration was begun with the minimum-energy solution, in other words the solution to (3) with $p = 2$. This is simply the inverse DFT of the vector matching Φx in the appropriate M places and zero elsewhere. This was done for 40000 iterations, by which point the ℓ^p norm had generally ceased to decrease significantly. The final value of ϵ was thusly 1.23×10^{-11}. The elapsed time was approximately 30 seconds on a 1.3-MHz laptop, running in MATLAB.

For each repetition, the reconstruction was deemed exact if the obtained minimizer u^* satisfied $\|u - x\|_\infty < 10^{-6}$. In each of several checked instances when this condition was satisfied, further iteration would produce a solution within less than 10^{-13} of x, an exact reconstruction by any reasonable numerical measure. The percentage of exact reconstructions was recorded for each K and M.

A pair of examples are in Figures 1 and 2. The sparsity is $K = 16$, and $M = 48$ measurements were made. The signal and measurement frequencies are the same for both examples. What differs is that $p = 1$ in the first case and $p = 0.5$ in the second case. When $p = 1$, the minimizer u^* differs substantially from the original signal x. The maximum discrepancy is $\|u^* - x\|_\infty = 1.08$, and the relative error is $\|u^* - x\|_2 / \|x\|_2 = 0.407$. The ℓ^1 norm of the minimizer is indeed lower than that of the signal, $\|u^*\|_1 = 22.7$ versus $\|x\|_1 = 23.4$. In the $p = 0.5$ case, the reconstruction is exact: $\|u^* - x\|_\infty = 7.61 \times 10^{-7}$ after 40000 iterations. Another 50000 iterations gave $\|u^* - x\|_\infty = 3.33 \times 10^{-14}$.

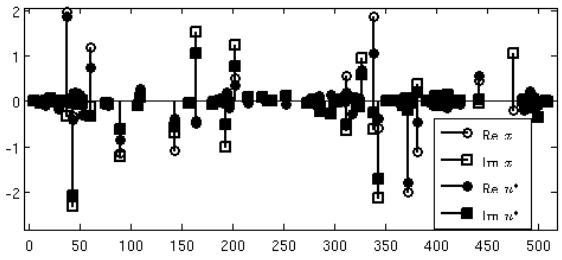


Fig. 1. The real and imaginary parts of a signal x (open markers) and the solution u^* (filled markers) to problem (2). The reconstruction is not even approximately x.

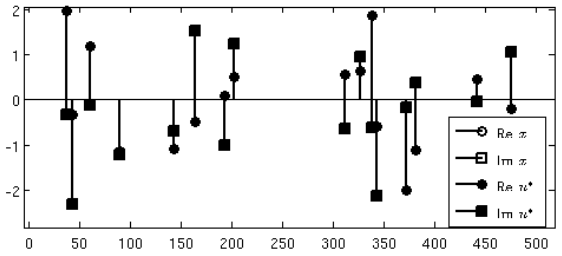


Fig. 2. Same as the previous figure, except u^* is the solution to problem (3) with $p = 0.5$. The reconstruction is exactly x.

Figure 3 shows the results of the experiments, in the format of [3]. The number of measurements M is on the vertical axis and the ratio K/M of sparsity to number of measurements is on the horizontal axis. (Only values of K/M up to $3/4$ are displayed.) The intensity gives the observed frequency of exact reconstruction, with white being 100% and black being 0%. Improvement over the $p = 1$ case is clear even for $p = 0.95$. For $p = 0.5$, exact reconstruction is obtained with a sparsity-to-measurement ratio nearly double that required when $p = 1$.

An additional experiment was conducted to provide an example of the result of varying the number of measurements with a fixed sparsity. We used $K = 16$ and values of M ranging from 24 to 80, for the same values of p used above. The results are in Figure 4. It is remarkable that even a value of p only slightly less than 1 gives exact reconstruction for significantly fewer measurements (12% fewer, in this example). Decreasing p further decreases the required number of measurements, but it appears that the smaller the value of p, the less improvement is seen for a given decrease in p.

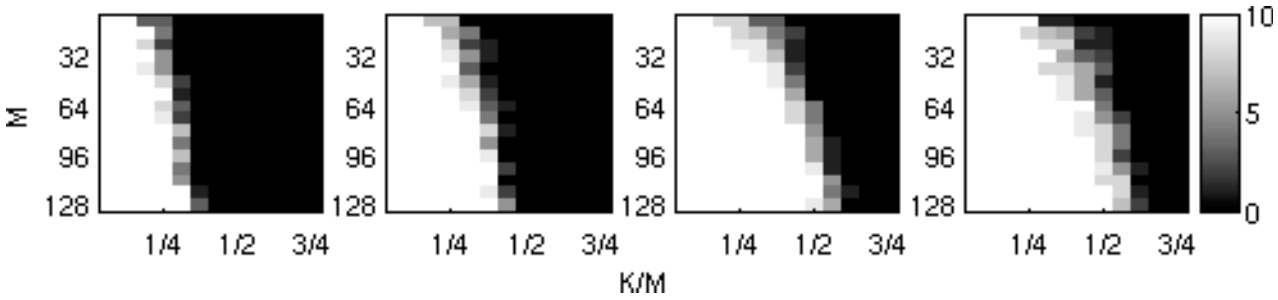


Fig. 3. Observed probabilities of exact reconstruction for different numbers of measurements (M) and sparsity-to-measurements ratio (K/M), for four values of p: $p = 1$ (left), $p = 0.95$ (second from left), $p = 0.75$ (second from right), and $p = 0.5$ (right). Compared with $p = 1$, exact reconstruction is obtained with larger values of K/M even for $p = 0.95$, and almost double the value of K/M for $p = 0.5$.

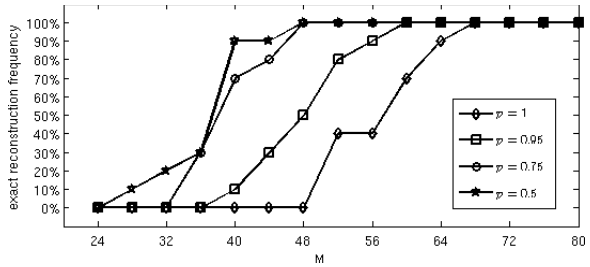


Fig. 4. Observed probabilities of exact reconstruction for a signal of sparsity K, for different numbers of measurements M and four values of p. Compared with $p = 1$, substantially fewer measurements are required for exact reconstruction when $p = 0.95$. Decreasing p further decreases the required value of M, but by less and less as p gets smaller.

III. Two-dimensional example

We present a larger-scale, two-dimensional example, designed as in [3], where the signal x is the 256×256 Shepp-Logan phantom[3] (Figure 5(a)). Here, it is the gradient of the signal that is sparse, with $\|\nabla x\|_0 = 3627$ (out of 65536 pixels). The compressed-sensing theory applies in this setting as well, as is well known [3]. The problem (3) becomes:

$$\min_u \|\nabla u\|_p, \quad \text{subject to} \quad \Phi u = \Phi x. \tag{4}$$

As in [3], the Fourier coefficients measured by Φ will not be at random frequencies, but instead along radial lines in frequency space (as in Figure 5(b)). By the Fourier slice theorem, sampling along a radial line in frequency space is equivalent to sampling (the Fourier transform of) the Radon transform along the angle determined by the line. The experiment is equivalent to attempting to reconstruct the phantom from a limited number of radiographic projections (or views).

We used gradient descent with projection as in Section II, but with ϵ and the step size chosen manually as the iteration proceeds. Roughly 10^5 iterations were required, taking about four hours. (Given that the signal was 128 times larger than in the one-dimensional examples and took very roughly 480 times longer, one can make the crude estimate of an $N \log N$ scaling.) Gradients were computed with simple forward differencing, with backward differencing used for divergences. In both cases, the boundary conditions were periodic.

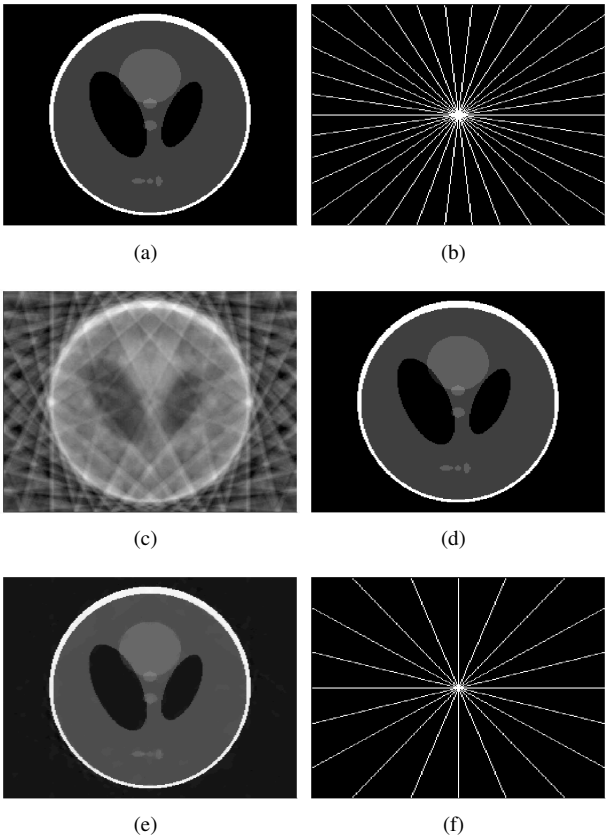
In [3], exact reconstruction is claimed for 22 views in the $p = 1$ case, for which there are $M = 5481$ Fourier coefficients measured. In fact, we find that 17 views suffice (or $M = 4257$; see Figure 5(d)), while 16 views do not ($M = 4017$). For $p = 0.5$, however, 10 views ($M = 2521$) are sufficient for exact reconstruction. Decreasing p further was not found to decrease the number of views required. That exact reconstruction can occur with a sparsity-to-number-of-measurements ratio exceeding 1 must surely rely on special properties of the phantom or of the sampling pattern, as this is not generally possible even with $p = 0$.

We also examine the effect of noise in the measurements. We perturb the measured Fourier coefficients by randomly chosen complex numbers, with standard-normally distributed real and maginary parts. Using $p = 0.5$ and 10 views, the reconstruction is a close approximation of the true phantom. A few specks are all that can be observed. The ℓ^∞ error is $\|u^* - x\|_\infty = 0.0938$, and the ℓ^2 relative error is $\|u^* - x\|_2 / \|x\|_2 = 0.0163$. With $p = 1$ and 17 views, a reasonable approximate phantom is obtained, with some blotchiness. The ℓ^∞ error is $\|u^* - x\|_\infty = 0.245$, and the ℓ^2 relative error is $\|u^* - x\|_2 / \|x\|_2 = 0.0505$.

IV. Conclusions

The ability to reconstruct signals from very few measurements is an important development in signal processing. In applications where data acquisition is expensive or difficult, compressive sensing can allow good results to be obtained in a manner that would once have been infeasible. In this note, we have seen

[3]The text of [3] gives conflicting implications about whether it is the 256×256 or 512×512 phantom that is used. Examining the figures confirms that it is the 256×256 phantom.



(a) (b) (c) (d) (e) (f)

Fig. 5. (a) The 256×256 Shepp-Logan phantom. (b) The white pixels show where the Fourier transform is sampled with 17 views. (c) The minimum-energy or backprojection reconstruction with 17 views is very poor. (d) The reconstruction from 17 views with $p = 1$ is exact. (e) The reconstruction from 16 views with $p = 1$. It is not quite equal to the phantom, and has a smaller ℓ^1 norm. (f) The Fourier transform sample pattern for 10 views.

that by using ℓ^p minimization with $p < 1$, fewer measurements are required than previously observed. The required reconstruction time is generally longer than with $p = 1$, but much less than with $p = 0$. Moreover, the algorithmic approach in this note is relatively naive; further research on the ℓ^p optimization problem should reduce the reconstruction time further. This will increase the number of applications in which the ℓ^p approach is worthwhile.

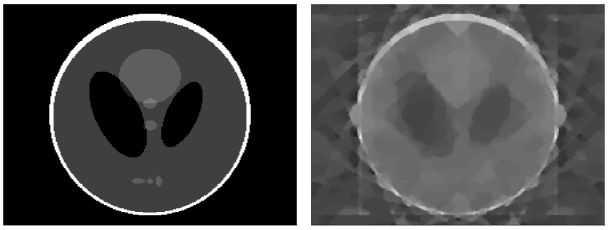
REFERENCES

[1] E. Candes and T. Tao, "Near optimal signal recovery from random projections: universal encoding strategies?," Tech. Rep. 04-70, UCLA Group in Computational and Applied Mathematics, December 2004.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, 2006.

[3] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, 2006.

[4] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1998.



(a) (b)

Fig. 6. (a) The reconstruction from 10 views with $p = 0.5$ is exact. (b) The reconstruction from 10 views with $p = 1$ is poor.

(a) (b)

Fig. 7. (a) The reconstruction from 17 noisy views with $p = 1$ is reasonable, but blotchy. (b) The reconstruction from 10 noisy views with $p = 0.5$ is good, having only a few specks.

[5] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, pp. 227–234, 1995.